Supplementary Files

Hepatic stellate cell activation and senescence induced by intrahepatic microbiota disturbances drive progression of liver cirrhosis towards Hepatocellular Carcinoma

Supplementary materials and methods

DNA extraction and bacterial 16S rRNA sequencing

Bacterial genomic DNA was extracted using QIAGEN DNA QIAamp DNA Mini Kit, according to the manufacturer's instructions. The 16S rDNA V4 region was amplified using the conserved primers 515F (5′-GTG CCA GCMGCC GCG GTAA-3′) and 806R (5′-GGA CTA CHVGGG TWT CTAAT-3′), and no template DNA reaction was used as a negative control. PCR products were monitored using the 2% agarose gel. PCR fragments were sequenced in an Illumina MiSeq platform (Illumina) using the 2x250 bp paired-end protocol yielding pair-end reads that overlap almost completely. 16S (variable region 4 [v4]) rRNA gene pipeline data incorporated phylogenetic and alignment-based approaches to maximize data resolution. The read pairs were demultiplexed based on unique molecular barcodes added via PCR during library generation, then merged using USEARCH v1.1 pipeline.

16S rRNA data analysis

Raw paired-end 16S rRNA reads (V4 region) were merged into consensus fragments by FLASH and subsequently filtered for quality (targeted error rate <

0.5%) and length (minimum 200bp). QIIME software 1.9 package was used to analyze sequences (Quantitative Insights Into Microbial Ecology, http://bio.cug.edu.cn/qiime /). Reads were clustered into operational taxonomic units (OTUs) at 97% identity (Vsearch). Alpha- and beta-diversity analysis were calculated using the relative abundance of OTUs in each sample. Differential abundance analysis of alpha diversity features of interest evaluated differences using the nonparametric difference test. Differential abundance analysis of taxonomic abundances evaluated differences using the negative binomial test (DESeq). The false discovery rate (FDR) was used to correct for multiple hypothesis testing. LEfSe was used for linear discriminant analysis. The 16S rRNA data were assessed the potential multiple biological pathways of the microbiota using PICRUSt, which provides proportional contributions of KEGG categories for each sample. Receiver Operator Characteristic (ROC) and Area Under Curve (AUC) analysis were performed using R, evaluating variable thresholds for relative abundance of taxa individually and in aggregate for classification of HCC with cirrhosis versus without cirrhosis status.

Tissue culture based and 16S rDNA PCR

Fresh HCC or normal liver tissue samples were homogenized under sterile conditions and cell fraction was pelleted. The supernatants were plated on Columbia agar plates and maintained under aerobic or anaerobic conditions overnight at 37°C. Microbiota colonies were selected and DNA was extracted for 16S rRNA sequencing.

For 16S rDNA PCR, bacterial DNA was extracted from frozen tissue samples in mouse model in sterile conditions using DNA QIAamp DNA Mini Kit (QIAGEN, Germany). 16S rDNA PCR was executed using the primers 515F–806R targeting the V4 region of bacteria and SM1-SM4 targeting *S.maltophilia*. The primers are listed in supplementary Table S2.

Fluorescence *in situ* Hybridization (FISH) and cell staining

Paraffin-embedded specimens were sectioned and then dewaxed. Proteinase K (50 μg/ml, Beyotime) solution was added to the slides and the slides were covered and incubated for 30 minutes at 37 °C. Next, the FISH probe EUB338 (5'-GCTGCCTCCCGTAGGAGT-3') or Stemal (5'-GTCGTCCAGTATCCACTGC-3') was added at 400 nM and incubated overnight in a humid environment at 37~45 °C. The samples were washed with 0.1 M Tris-HCl (pH 7.4) and then blocked with 0.1 M pH 7.4 (2% FBS, 0.3% Triton X) for 30 minutes at room temperature. The slides were incubated with anti-human CD45 antibodies for an hour followed by fluorophore-conjugated secondary antibodies for another hour. The slides were mounted with DAPI-Antifade Solution in the dark place, incubated for 10 minutes and observed under a fluorescence microscope with the proper filter sets.

Bacterial Strains and Culture Conditions

Standard strains of *S. maltophilia* were purchased from BeNa Culture Collection, (Beijing, China). The bacteria were cultured on Colombian blood agar plates. Single colonies were selected and cultured in brain-hearth infusion broth (BHI)

with agitation at 37 °C overnight. Some of the bacteria were frozen in BHI with 10% glycerin at - 80 °C, and the rest of the bacteria were was stored at 4 °C for cell stimulation.

Cell culture

The LX-2 cell line was purchased from American Type Culture Collection (ATCC, Virginia, USA) and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin and 100 μg/ml streptomycin.

Western blot

Liver tissues or cells were lysed in radioImmunoprecipitation Assay (RIPA) buffer containing a protease inhibitor cocktail (Thermo Fisher). The lysates were centrifuged at 12,000 g for 30 minutes at 4 °C. The supernatants were collected for further analysis. The total protein concentration was measured using a BCA assay kit (Beyotime). Equal amounts of total proteins were separated with a 12% SDS-PAGE gels and then transferred onto polyvinylidene fluoride (PVDF) membranes using a wet transfer device. After blocking with 5% nonfat milk at room temperature for 1 hour, the membranes were incubated with primary antibodies overnight at 4 °C. The membranes were then incubated with the appropriate HRP-conjugated secondary antibodies for 1 hour at room temperature followed by incubation with an electrochemiluminescence reagent (Beyotime). The antibodies are listed in supplementary Table S3. All the values were normalized to GAPDH.

Immunohistochemistry (IHC)

Briefly, formalin-fixed paraffin-embedded specimens were processed into 4-μm thick sections. The sections were deparaffinized and rehydrated, the endogenous peroxidase was blocked and the antigens were retrieved. The antibodies are listed in Table S3. After incubation overnight at 4 °C with primary antibodies, the sections were then incubated with HRP-conjugated secondary antibodies at room temperature for 1 hour, followed by color development with the 3,3'-diaminobenzidine (DAB) substrate. The nuclei were counterstained with hematoxylin. The percentage of stained cells was calculated by dividing the number of stained cells by the total number of hematoxylin-stained nuclei which were visualized in ten random high-power fields (400×). All the sections were evaluated by two independent observers.

Immunofluorescence (IF)

The antibodies are listed in supplementary Table S3 and the nuclei were counterstained with DAPI. Images were obtained using an OLYMPUS BX53 (Olympus, Japan) fluorescence microscope.

Mouse Primary Hepatic Stellate Cell Culture

Mouse HSCs were isolated from adult BALB/c mice by retrograde pronase-collagenase liver perfusion and density centrifugation. Perfusion through inferior vena cava was initiated with 15 mL of 190 mg/L EGTA solution. Portal vein was cut and vena cava clamped above the diaphragm to ensure correct liver perfusion. Then, perfusion continued with 25 ml of 0.4 mg/ml pronase

solution and, finally, 35 ml of 0.1 U/ml collagenase solution. The liver was removed, minced, and incubated (10−20 minutes, 40℃ under agitation) in pronase/collagenase solution containing 1% (vol/vol) DNase. The resulting dispersed cell suspension was filtered through 70-μm cell strainer and centrifuged (50g, 5 minutes). HSCs were separated by Histodenz density gradient centrifugation (Sigma, St Louis, MO) and culture-activated on 6-well Nunc plates (Thermo-Fisher, Waltham, MA) containing 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, and 2 mM l-glutamine. HSCs were confirmed by α-SMA staining and quantitative PCR analysis.

Quantitative PCR

RNA was detected using qRT-PCR. The relevant primers are listed in supplementary Table S2. Briefly, Total RNA was extracted using the TRIzol reagent (Takara, Japan). RNA was reverse transcribed with the HiScript®Q Select RT SuperMix for qPCR (+gDNA wiper) Kit (Vazyme, Nanjing, China), Real-time PCR was performed with the AceQ® qPCR SYBR Green Master Mix (High ROX Premixed) Kit (Vazyme, Nanjing, China). The results were standardized to control values of glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Sirius red staining

A Sirius red staining kit (Cat. G1471; Solarbio) was used for the staining. Briefly, after dewatering and embedding, the tissues were cut into 6-μm-thick sections and dewaxed to water. The sections were stained with Weigert ferricin staining

solution for 20 minutes and then washed with tap water for 5 minutes, followed by washed with distilled water. The slides were then stained with a drop of Sirius red dye and incubated for 1hour before slightly rinsing with the tap water to remove the dye on the surface of the sections.

β-gal staining

β-gal staining was performed using a β-gal staining kit (Cat. C0602; Beyotime). Briefly, the paraffin sections were dewaxed and hydrated and fixed with the kit at room temperature for 15 minutes. After washing with PBS, the dye solution was added and incubated overnight at 37 °C. For adherent cells cultured in 6-well plates, 1 ml staining fixative buffer was added to each well after the medium was removed, and the plates were incubated at room temperature for 15 minutes. After washing with PBS, 1 ml of dyeing solution was added to each well and incubated overnight at 37 °C. The next day, the cells were observed under a microscope.

Hepatic Fibrosis in a Mouse Model

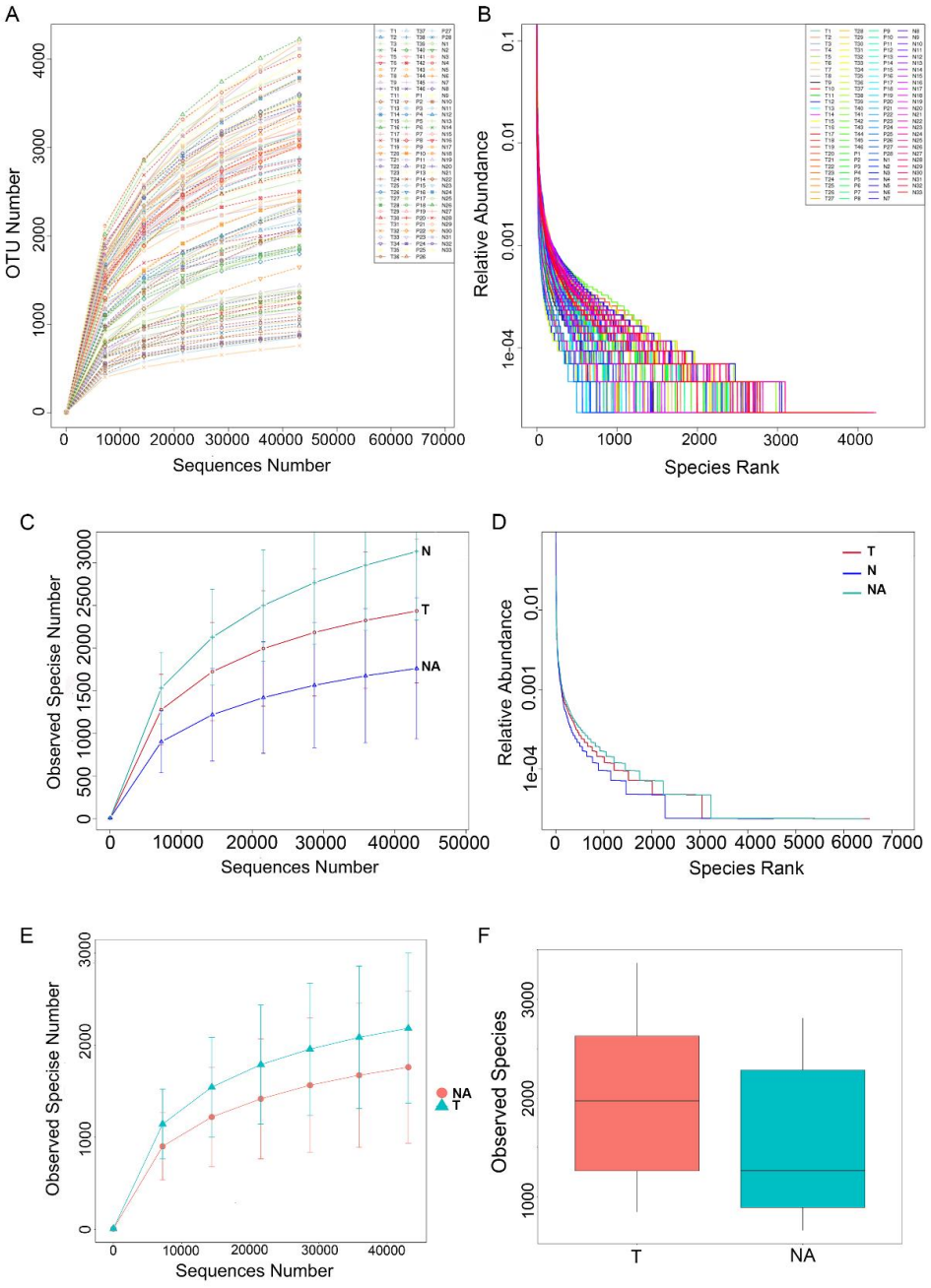
Mouse strains used in this study included NLRP3-deficient mice and C57BL/6 wildtype (WT) mice (provided by Animal Core Facility of Nanjing Medical University). Some mice were injected intraperitoneally with 20% CCl4, 0.1ml/10g, twice a week for 8 weeks, to make liver fibrosis model. Some six- to eight-week-old mice were infected through the intraperitoneal route with 5×10^5 CFU of S.maltophilia. Mice were born from breeding pairs that were housed under controlled temperature (22±2 °C) and artificial light under a 12-h cycle

period. Mice were kept under specific pathogen-free conditions in positive-pressure cabinets and provided with sterile food and water ad libitum.

Statistical analysis

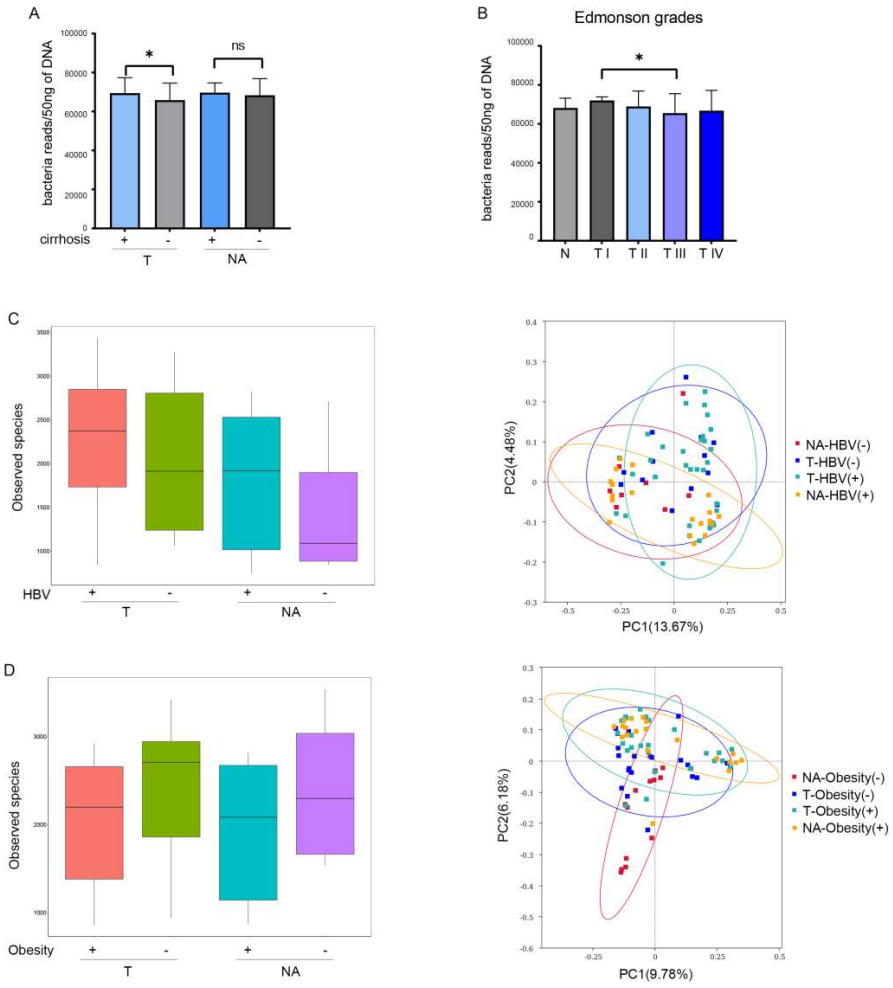
The clinical and sequencing data are presented as medians (first quartile to the third quartile). The statistical significance of differences among groups of specimens was tested with the one-way analysis of variance of Friedmans test, Dunn posthoc test, or Wilcoxon rank-sum test (according to data distribution). Weighted UniFrac distance metric and principal component analysis were used to perform beta-diversity analysis to observe the overall changes in the species compositions of tissues. Statistical tests were performed using R (V.2.15.3). GraphPad Prism 8 (La Jolla, CA, USA) was used to create graphs. *P* values below 0.05 were considered statistically significant.

Supplementary figures



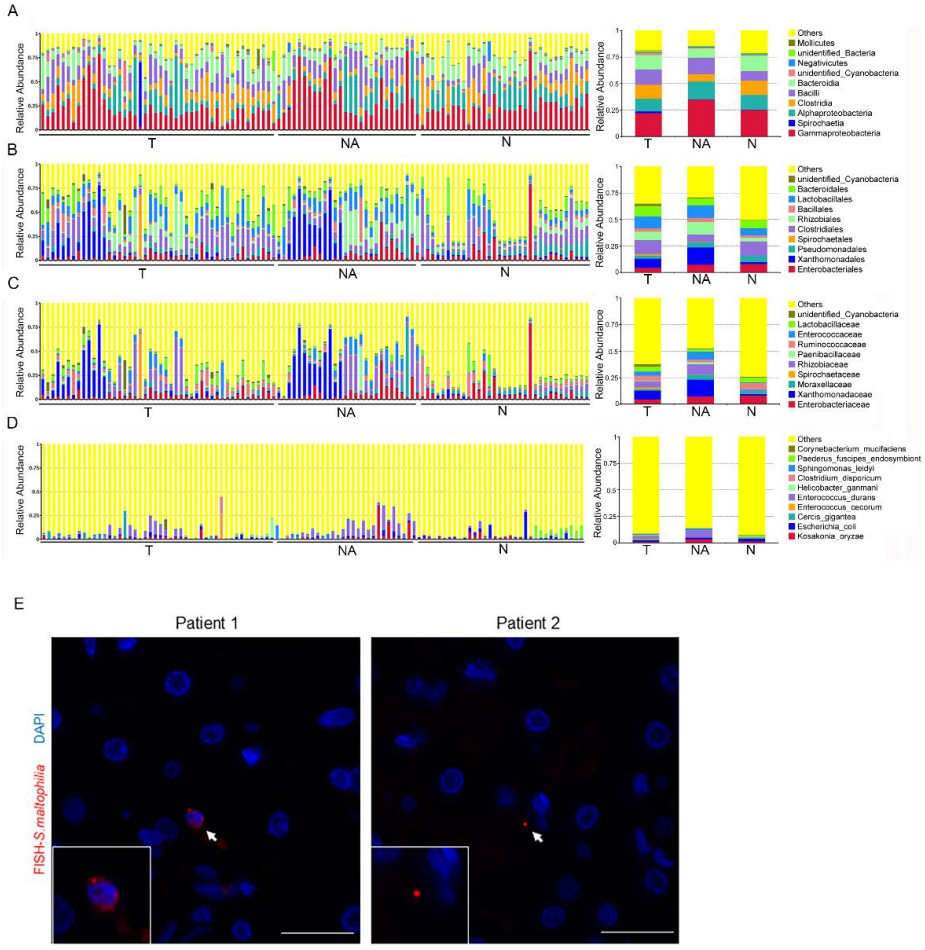
Supplementary Figure S1. Tumor microbiome dysbiosis in HCC. (A-B) The rarefaction curve (A) and the rank abundance curve (B) for each individual specimen are shown. (C) Rarefaction curves were used to estimate the

richness (at a 97% level of similarity) of the microbiota among T, NA and N groups. The vertical axis shows the number of OTUs expected after sampling the number of tags or sequences shown on the horizontal axis. (D) Relative abundance curves of the bacterial OTUs derived from T, NA and N groups. (E) The rarefaction curve of 28 paired T and NA sequencing datasets. Tumor tissues revealed a relatively higher diverse compared to matched normal adjacent samples of HCC patients (n=28). (F) α diversity analysis of 28 paired T and NA tissues. The majority of the tumoral samples saturated around 1980 species and around 1254 species for the paired normal adjacent samples (P = 0.038). Data are presented as means ± SEM. Unpaired Student's t test was performed for (F).

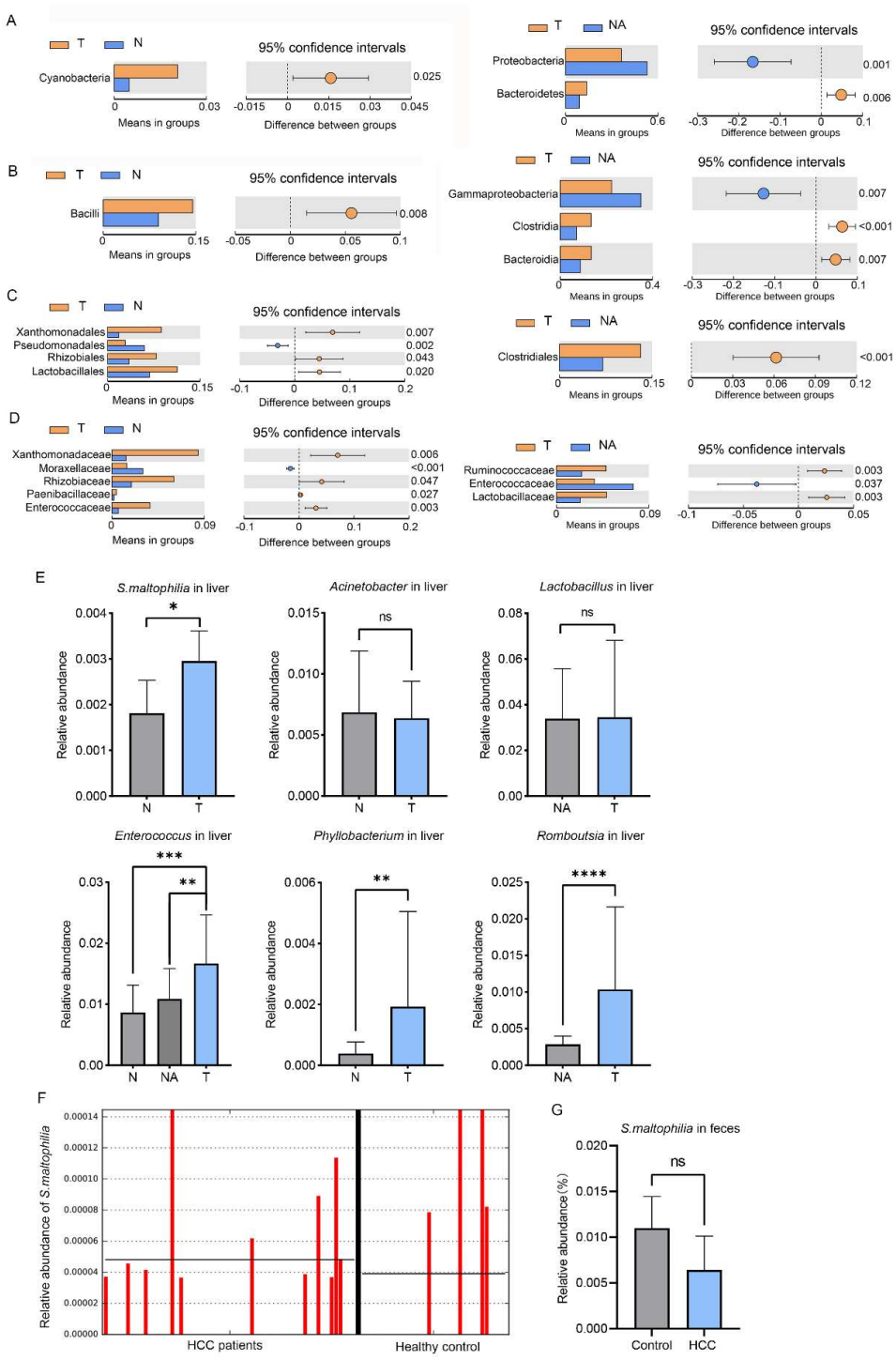


Supplementary Figure S2. The influences of cirrhosis, progression of HCC, HBV, alcohol over consumption and obesity on intrahepatic microbes in HCC**.** (A-B) The bacterial reads in human livers were assessed by bacterial 16S rDNA sequencing. Levels of bacterial DNA in tumors with cirrhosis were significantly higher than those without cirrhosis in HCC patients, $P < 0.05$ (A). With the progression of HCC, bacterial DNA was observed decreased, $P < 0.05$ (B). (C-D) α (left) and β (right) diversity analysis with or without HBV infection (C), obesity (D) in T and NA tissue. None of these data has significant *P* value. Data

are presented as means ± SEM. ns, no statistical significance. **P* < 0.05. Unpaired Student's *t* test was performed for (A, C, and D) and one-way ANOVA test for (B).

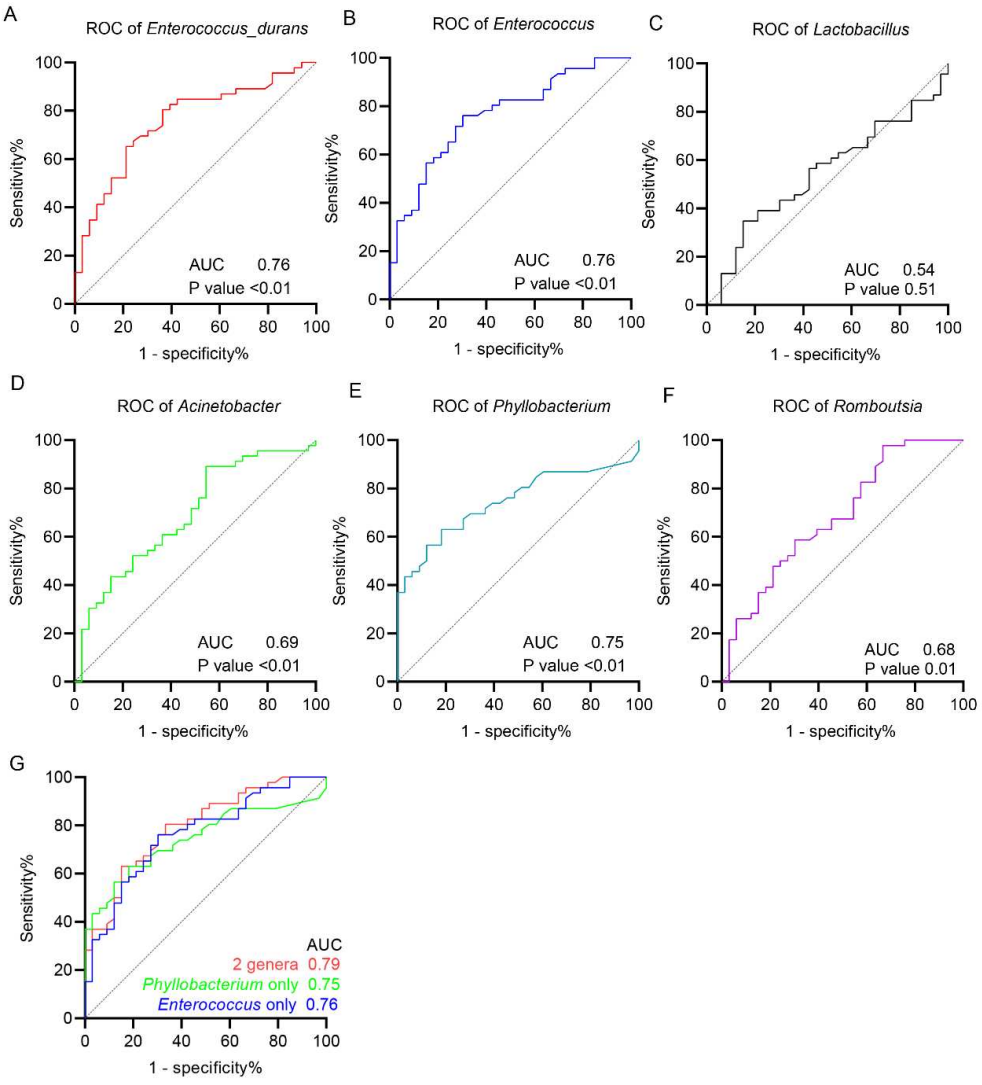


Supplementary Figure S3. Taxonomic composition between T, NA and N groups. (A-D) Bar plots displaying taxonomic composition in both cohorts at Class (A), Order (B), Family (C) and Species (D), respectively. (E) Immunofluorescence with a specific FISH probe for ***S. maltophilia*** only (Stemal). Liver sections were stained with FISH probe (red) and DAPI (blue). Scale bar indicates 50 μm.

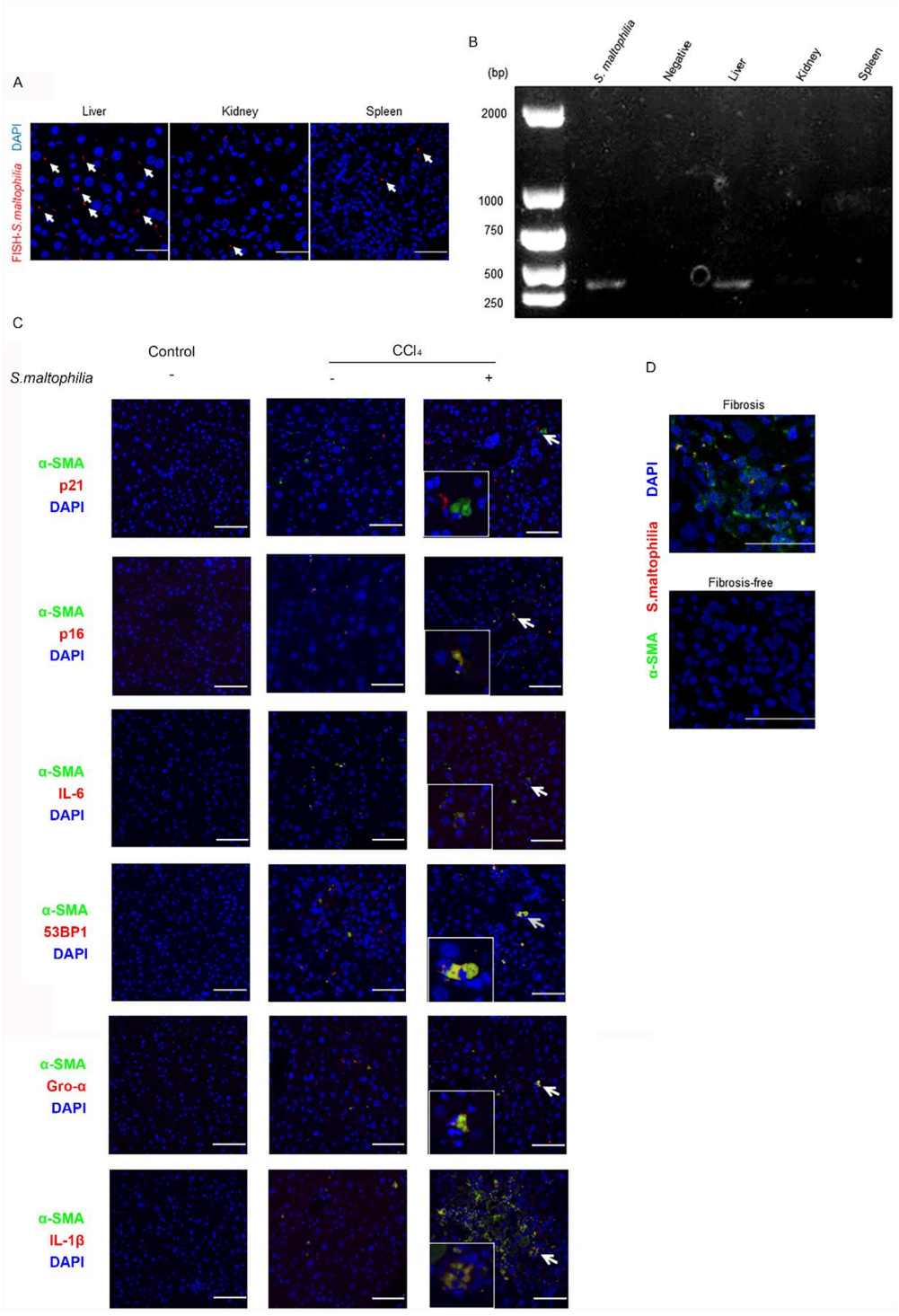


Supplementary Figure S4. Differences in the classification levels between the groups. Mann–Whitney U-test of different bacteria between T vs N and T vs NA in the classification levels. From A to D is (A) Phylum, (B) Class, (C) Order, (D)

Family. (E) Mann–Whitney U-test of different bacteria in independent cohort. (F) Relative abundance of S.maltophilia from gut microbiota in HCC. No significant difference was observed between HCC and control groups (G). Data are presented as means ± SEM. ns, no statistical significance. Unpaired Student's *t* test was performed for (G) and one-way ANOVA test for (F).



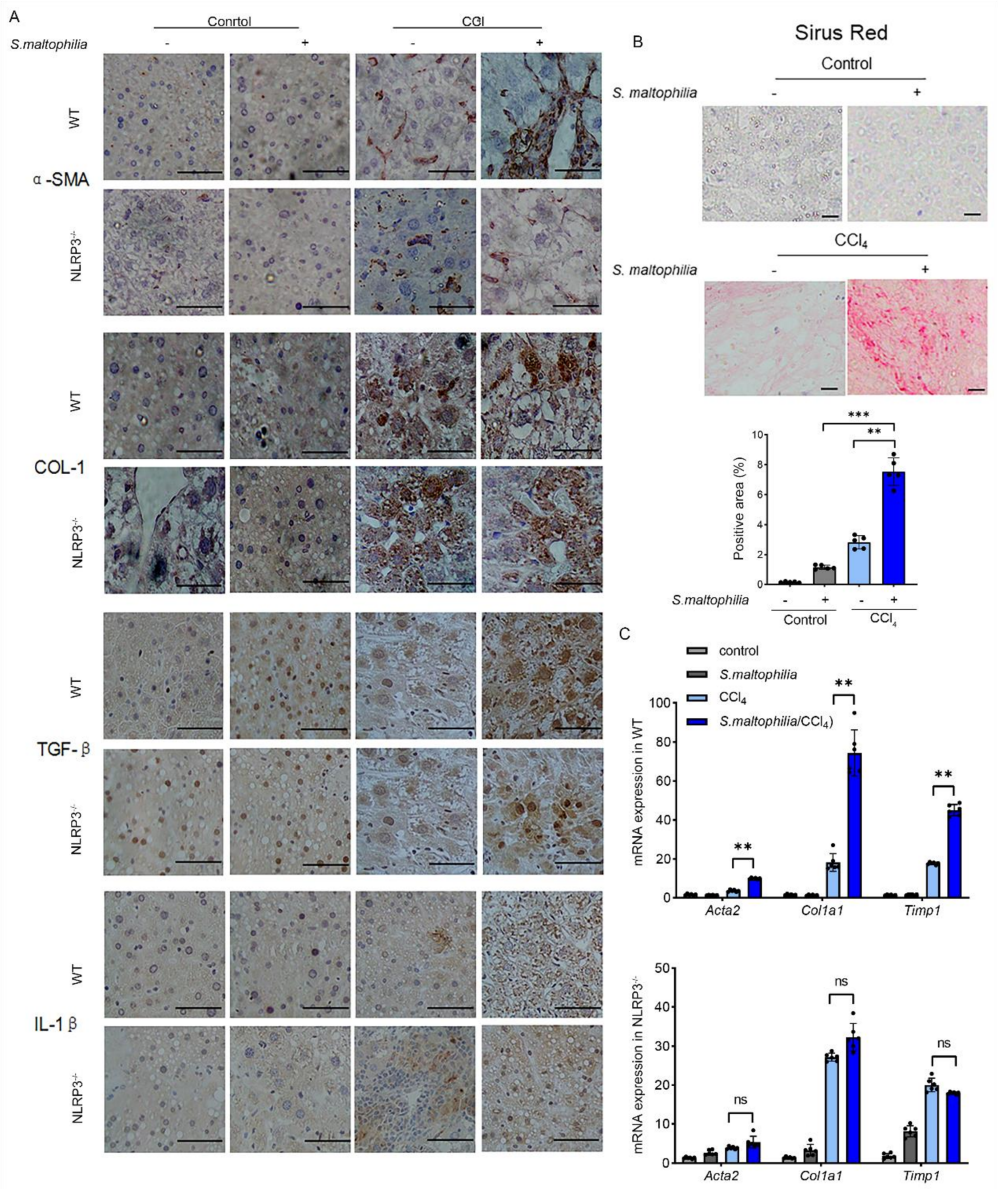
Supplementary Figure S5. The ROC curve of significant bacteria. The ROC curves of other different microbiota except for *S. maltophilia* were shown in (A-G). If *S. maltophilia* was excluded, the maximum combined AUC value of the remaining microbiota was 0.79 ($P < 0.0001$).



Supplementary Figure S6. *S. maltophilia* induces the senescence of hepatic stellate cells and fibrosis in mice liver. (A) Immunofluorescence with a specific FISH probe for *S. maltophilia* only (Stemal) in mice liver, kidney and spleen.

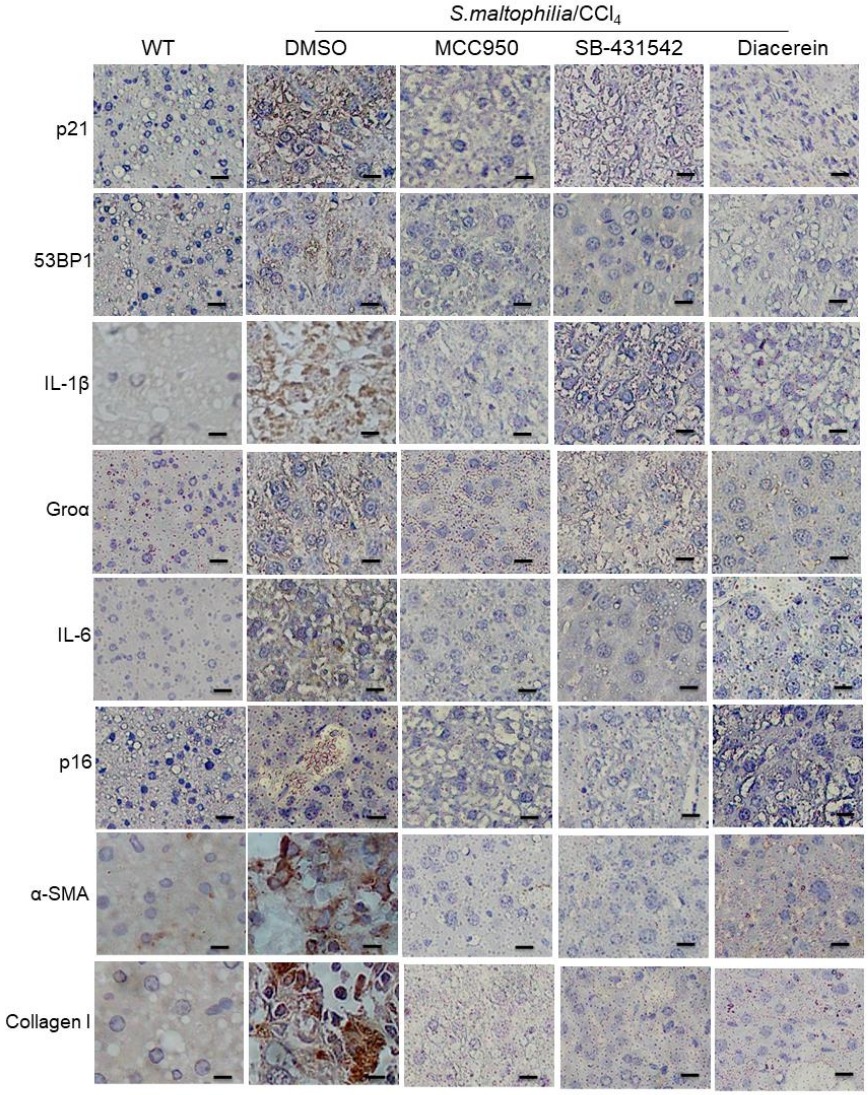
Liu B, *et al. J Immunother Cancer* 2022; 10:e003069. doi: 10.1136/jitc-2021-003069

Sections were stained with FISH probe (red) and DAPI (blue). Scale bar indicates 50 μm. (B) PCR of *S. maltophilia* on liver, kidney and spleen in mice after intraperitoneal injection with CCl4 and *S. maltophilia*, *S. maltophilia* and *L. reuteri* as positive and negative controls,respectively. (C) Representative of immunofluorescence images of HSCs of liver sections in mice with or without treatment with *S. maltophilia*. HSCs were visualized by α-smooth muscle actin staining (α-SMA; green), and the cell nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI; blue). Arrowheads indicate α-SMA-expressing cells that were positive for SASPs including p21, 53BP1, IL-1β, Groα, IL-6 and p16 (red). Scale bar indicates 50 μm. (D) Representative of immunofluorescence images in mice liver with α-SMA (green) and *S. maltophilia* specific probe (red). Scale bar indicates 50 μm.



Supplementary Figure S7. *S. maltophilia* induces fibrosis in mice liver. (A) Immunohistochemistry of Collagen I, TGF-β, α-SMA and IL-1β in WT or NLRP3-deficient mice liver with or without treatment with *S. maltophilia*, Scale bar, 50 μm. (B) Fibrillar collagen deposition was evaluated by Sirius red staining in mice livers after different stimulations. (Each group had 5 mice.) Scale bar, 100 μm. ** *P*<0.01. (C) The levels of *Col1a1*, *Acta2* and *Timp1* mRNA were measured

by qPCR in WT or NLRP3-deficient mice liver with or without treatment with *S. maltophilia.* (Each group had 5 mice.) Data are presented as means ± SEM of four independent experiments (B and C). ns, no statistical significance. **P* < 0.05, ***P* < 0.01, and ****P* < 0.001. Two-way ANOVA was performed for (B), and one-way ANOVA test for (C).



Supplementary Figure S8. NLRP3, IL-1β and TGF-β inhibitors attenuated fibrosis induced by *S. maltophilia* in mice. (A) The expressions of α-SMA, Collagen I and SASPs including p21, 53BP1, IL-1β, Groα, IL-6 and p16 were detected by immunohistochemistry in mice liver sections after treatment with *S. maltophilia* or together with TLR-4 inhibitor TAK-242, NF-κB inhibitor PS-341 or/and the NLRP3 inhibitor MCC950. Scale bar indicates 100 μm.

Supplementary Tables

Table S1-1. Summary of clinical characteristics of HCC patients.

Characteristics	Patients (n = 46)	Independent cohort patients （n=37）
Age (means, min, max)	59.84 (40, 80)	61.57(39, 79)
Gender (Female/Male)	14/32	9/24
Complications		
Hypertension	13	9
Diabetes	8	4
Cirrhosis	20	20
Hepatitis B	36	27
Tumor Stage		
I	3	1
II	12	9
III	25	25
IV	6	2
Tumor Volume (means, min, max)	285.20 (6,1800)	262.39(1.5, 2170)
Tumor Screening Indicators		
AFP (I/II/III/IV)	23/5/6/12	12/9/6/10
CEA (means, min, max)	7.85 (0.58, 254.20)	6.75(1.01, 10.45)
CA199 (means, min, max)	34.24 (0.86, 342.20)	27.50(0.91, 79.53)
Plasma Biochemical Index		
TB (means, min, max)	26.28 (7.2, 396.1)	13.9(7.1, 26.3)
ALT (means, min, max)	79.87 (7.4, 859.3)	103.9(7.6,524.9)
AST (means, min, max)	77.53 (14.9, 629.4)	80.6(12.7, 585.5)
GGT (means, min, max)	110.58 (16.0, 868.8)	68.8(18.3, 296.2)
Sample collection		
Tumor	46	33
Tumor stage (I/II/III/IV)	3/12/25/6	1/8/22/2
normal adjacent	28	27
Tumor stage (I/II/III/IV)	0/5/20/3	0/6/20/1

Table S1-2. Summary of clinical characteristics of hepatic hemangioma patients.

Characteristics	Patients (n = 33)	Independent cohort patients （n=14）
Age (means,min,max)	55.78(37,70)	51.5(38, 63)
Gender(Female/Male)	14/19	4/10
Complications		
Hypertension	5	1
Diabetes	2	1
Cirrhosis	2	3
Hepatitis B	4	1
Hemangioma / Hepatolith	18/15	
Tumor Screening Indicators		
AFP(I/II/III/IV)	31/2/0/0	13/1/0/0
CEA (means,min,max)	3.43(0.26, 10.88)	2.71(0.74, 6.88)
CA199 (means,min,max)	17.81(5.97, 24.50)	18.36(0.60, 29.53)
Plasma Biochemical Index		
TB (means,min,max)	21.36(6.9, 125.5)	20.00(9.0, 26.3)
ALT (means,min,max)	43.03(10.6, 139.9)	59.94(7.4, 109.3)
AST (means,min,max)	43.02(15.7, 154.0)	43.78(19.2, 104.7)
GGT (means,min,max)	47.28(14.0, 124.4)	62.07(16.0, 175.6)

The clinical information of 79 patients in the first sequencing and 51 patients in the independent cohort is shown in the table. Age, tumor volumes, tumor screening indicators, and plasma biochemical indexes are presented as the means, with the minimum and maximum values in parentheses.

Supplementary Table S2

Primer's list.

mRNA qPCR primers		
Gene	**Forward (5′-3′)**	**Reverse (5′-3′)**
GAPDH (mouse)	CAACTACATGGTCTACATGTTC	CACCAGTAGACTCCACGAC
IL-1β (mouse)	GGACCCATATGAGCTGAAAGCT	TGTCGTTGCTTGGTTCTCCTT
IL-6 (mouse)	AGAAGGAGTGGCTAAGGACCAA	AACGCACTAGGTTTGCCGAGTA
Gro-α (mouse)	GCTGGGATTCACCTCAAGAA	AGGTGCCATCAGAGCAGTCT
HMGA1 (mouse)	GGCACTGAGAAGCGGGGCCG	CCCTTGTTTTTTGCTTCCCTT
Acta2 (human)	ACTGAGCGTGGCTATTCCTCCGTT	GCAGTGGCCATCTCATTTTCA
Tgfb1 (human)	CGACTACTACGCCAAGGA	GAGAGCAACACGGGTTCA
Timp1 (human)	GCACATCACTACCTGCAGTC	GAAACAAGCCCACGATTTAG
Col1a1 (human)	GGAACACCTCGCTCTCCA	GGGATTCCCTGGACCTAAAG
GAPDH (human)	GCCCAATACGACCAAATCC	AGCCACATCGCTCAGACAC
Bacterial primer		
Name	**Forward (5′-3′)**	**Reverse (5′-3′)**
16S rRNA primer	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT
***S.maltophilia* primer**	CAGCCTGCGAAAAGTA	TTAAGCTTGCCACGAACAG

Supplementary Table S3

Antibodies and inhibitors list.

Antibodies

Name	Manufacturers	Catalog number
Collagen I	Bioworld	BS1530
TGF-β	Abcam	ab92486
α-SMA	CST	19245
GAPDH	Beyotime	AF1186
IL-1β	Abcam	ab2105
CD45	Santa Cruz Biotechnology	sc-1178
α-SMA	CST	48938
Gro-α	Proteintech	12335-1-AP
53BP1	CST	88439
p21	Abcam	2947
p16	CST	ab51243
Alexa-Fluor 488-conjugated anti-mouse	CST	4408
Alexa-Fluor 647-conjugated anti-rabbit	CST	4414
P-IKK α/β	CST	2697
Caspase-1	CST	4199
NF-κB p65	CST	8242T
TLR4	Abcam	ab13556

Inhibitors

Name	Manufacturers	Catalog number
MCC950	MCE	HY-12815A
PS-341	MCE	HY-10227
TAK-242	MCE	HY-11109
Diacerein	MCE	HY-N0283
SB-431542	MCE	HY-10431

Table S4. Comparison of differential bacteria in theT, NA, and N groups.

Taxa	Name & Abundance ranking	Avg (T)	Avg (NA)	Avg (N)	*p*. value
phylum	*Proteobacteria (1, 2, 1)*	0.362 (0.189)	0.529 (0.197)	0.423 (0.157)	<0.001 (**T vs NA**)
	Bacteroidetes (3, 8, 6)	0.136 (0.090)	0.088 (0.057)	0.150 (0.063)	0.006 (**T vs NA**)
	Cyanobacteria (6, 10, 9)	0.021 (0.046)	0.008 (0.007)	0.005 (0.003)	0.025 (**T vs N**)
class	*Gammaproteobacteria (1, 1, 1)*	0.222(0.150)	0.349 (0.206)	0.251(0.128)	0.007 (**T vs NA**)
	Bacilli (2, 3, 5)	0.145 (0.099)	0.156 (0.115)	0.090 (0.083)	0.008 (**T vs N**)
	Bacteroidia (3, 4, 2)	0.135 (0.090)	0.087 (0.057)	0.147 (0.062)	0.007 (**T vs NA**)
	Clostridia (4, 5, 4)	0.134 (0.094)	0.070 (0.044)	0.135 (0.142)	<0.001 (**T vs NA**)
order	*Clostridiales (1, 5, 1)*	0.132 (0.090)	0.070 (0.044)	0.135 (0.142)	<0.001 (**T vs NA**)
	Lactobacillales (2, 3, 4)	0.113 (0.088)	0.119 (0.080)	0.068 (0.080)	0.020 (**T vs N**)
	Xanthomonadales (4, 1, 8)	0.087 (0.164)	0.161 (0.225)	0.018 (0.017)	0.007 (**T vs N**)
	Rhizobiales (5, 2, 6)	0.079 (0.141)	0.120 (0.147)	0.034 (0.030)	0.043 (**T vs N**)
	Pseudomonadales (8, 7, 5)	0.028 (0.022)	0.052 (0.137)	0.059 (0.050)	0.002 (**T vs N**)
family	*Xanthomonadaceae (1, 1, 6)*	0.084 (0164)	0.158 (0.227)	0.014 (0.015)	0.006 (**T vs N**)
	Moraxellaceae (9, 5, 4)	0.014 (0.012)	0.044 (0.138)	0.030 (0.017)	<0.001 (**T vs N**)
	Rhizobiaceae (2, 2, 5)	0.060 (0.034)	0.100 (0.145)	0.019 (0.025)	0.047 (**T vs N**)
	Paenibacillaceae (10, 8, 9)	0.005 (0.007)	0.020(0.087)	0.002 (0.002)	0.027 (**T vs N**)
	Enterococcaceae (6, 3, 7)	0.037 (0.065)	0.075 (0.079)	0.006 (0.013)	0.003 (**T vs N**)
					0.037 (**T vs NA**)
	Ruminococcaceae (4,6,2)	0.048 (0.048)	0.025 (0.017)	0.053 (0.080)	0.003 (**T vs NA**)
	Lactobacillaceae (3, 7, 3)	0.049 (0.049)	0.023 (0.022)	0.047 (0.072)	0.003 (**T vs NA**)

genus	*Stenotrophomonas (1, 1, 5)*	0.077 (0.166)	0.156 (0.225)	0.009 (0.015)	0.008 (**T vs N**)
	Acinetobacter (7, 4, 2)	0.012 (0.010)	0.043 (0.138)	0.019 (0.013)	0.010 (**T vs N**)
	Phyllobacterium (2, 2, 9)	0.056 (0.134)	0.100 (0.144)	<0.001 (<0.001)	0.008 (**T vs N**)
	Enterococcus (4, 3, 6)	0.037 (0.065)	0.074 (0.078)	0.006 (0.013)	0.003 (**T vs N**)
					0.037 (**T vs NA**)
	Lactobacillus (3, 6, 1)	0.048 (0.049)	0.023 (0.021)	0.047 (0.072)	0.003 (**T vs NA**)
	Romboutsia (9, 9, 4)	0.008 (0.013)	0.003 (0.003)	0.012 (0.050)	0.015 (**T vs NA**)
species	*Enterococcus_durans (1, 1, 4)*	0.025 (0.039)	0.066 (0.071)	0.005 (0.012)	0.002 (**T vs N**)
	Sphingomonas_leidyi (6, 3, 5)	0.006 (0.009)	0.017 (0.025)	0.002 (0.002)	0.003 (**T vs N**)

The microbiota in the T, NA, and N groups were compared at the phylum, class, order, family, genus, and species levels, as shown in the first column. The numbers in brackets in the second column represent the abundance rank in the T, NA, and N groups, respectively. The numbers in brackets in the third to fifth columns represent standard deviations. All the data were analyzed by Wilcoxon rank test.

Table S5: Function prediction in T and N

Taxa	avg(Tumor)	sd(Tumor)	avg(Normal)	sd(Normal)	p.value	q.values	interval lower	interval upper
Unclassified; Metabolism	0.025638	0.00099	0.026541	0.001461	0.003306	0.018941	-0.00149126	-0.000314464
Metabolism; Metabolism of Terpenoids and Polyketides	0.018823	0.001355	0.019791	0.002026	0.020523	0.050397	-0.001781755	-0.000155235
Metabolism; Biosynthesis of Other Secondary Metabolites	0.009471	0.000684	0.00907	0.000727	0.015778	0.050397	7.79E-05	0.000724089
Environmental Information Processing; Signaling Molecules and Interaction	0.001843	0.000218	0.001738	0.000203	0.032769	0.07041	8.77E-06	0.000199439
Human Diseases; Cancers	0.001573	0.000237	0.001329	0.000437	0.00269	0.023162	-0.000413296	-7.61E-05
Organismal Systems; Environmental Adaptation	0.001466	0.000126	0.001405	9.93E-05	0.019198	0.050397	1.02E-05	0.000111228
Human Diseases; Metabolic Diseases	0.000877	8.87E-05	0.000819	7.73E-05	0.002691	0.018941	2.09E-05	9.57E-05
Organismal Systems; Excretory System	0.000298	8.60E-05	0.000331	5.54E-05	0.042041	0.072266	-6.46E-05	-1.22E-06
Human Diseases; Cardiovascular Diseases	0.000161	8.64E-05	0.000217	0.000132	0.037869	0.072266	-0.000108715	-3.26E-06

Table S6: Function prediction in T and NA

Taxa	avg(Tumor)	sd(Tumor)	avg(NA)	sd(NA)	p.value	q.values	interval lower	interval upper
Metabolism; Energy Metabolism	0.056624	0.004375	0.054141	0.002495	0.002709	0.020721	0.000889577	0.00407574
Metabolism; Metabolism of Cofactors and Vitamins	0.041072	0.002291	0.039982	0.001838	0.027877	0.038023	0.000122111	0.00205677
Metabolism; Nucleotide Metabolism	0.034055	0.00286	0.032382	0.00335	0.032412	0.038023	0.000146015	0.003201465
Metabolism; Xenobiotics Biodegradation and Metabolism	0.029681	0.006407	0.033193	0.007055	0.036217	0.038023	-0.006789415	-0.000234184
Unclassified; Genetic Information Processing	0.024051	0.00203	0.022925	0.002074	0.026232	0.038023	0.000138043	0.00211358
Metabolism; Metabolism of Other Amino Acids	0.018153	0.001525	0.01914	0.001234	0.003332	0.020721	-0.001633081	-0.000339796
Human Diseases; Neurodegenerative Diseases	0.00273	0.000653	0.003129	0.00058	0.008067	0.033444	-0.000691006	-0.000107724
Human Diseases; Metabolic Diseases	0.000877	8.87E-05	0.000829	9.67E-05	0.036687	0.038023	3.09E-06	9.33E-05
Organismal Systems; Immune System	0.000568	0.000101	0.000525	7.67E-05	0.044433	0.042508	1.09E-06	8.41E-05